Entropy Based Estimation of Risk Neutral Density Dynamics

Ana Fernandes[2], Rui Pascoal[1], and Ana M. Monteiro[1(✉)]

[1] Univ. Coimbra, Faculty of Economics, Av Dias da Silva 165, 3004-512 Coimbra, Portugal
ruiapsp@fe.uc.pt, ammm@fe.uc.pt

[2] Univ. Coimbra, CeBER, Faculty of Economics, Av Dias da Silva 165, 3004-512 Coimbra, Portugal
anasofiapfernandes@hotmail.com

Abstract. Derivative contracts constitute a fundamental component of financial markets, providing valuable insights into traders' forward-looking expectations. In particular, option contracts enable to retrieve the risk-neutral density (RND) of the underlying asset price. The maximum entropy approach offers a flexible framework for estimating the RND, especially in the presence of incomplete market information. Peter Buchen and Michael Kelly (1996), along with Leonidas Rompolis (2010), were among the first to apply the maximum entropy principle to RND estimation.

This study proposes a relative measure to capture the day-to-day evolution of changes in the RND by jointly employing maximum entropy and minimum cross-entropy methods. Moments based Rompolis's methodology is implemented. To address the ill-conditioning problem of the Buchen and Kelly approach we propose a strike selection procedure based on open interest data. The methodology is applied to both simulated data and market data from Microsoft, as well as the S&P 500 and VIX indices. Concerning market data, the proposed relative measure allows us to conclude that daily changes in the risk-neutral density estimates are not significant. (This study results from a master's dissertation archived in the Scientific Repository of the University of Coimbra [16].)

Keywords: Entropy · density estimation · financial derivative contracts · optimization

1 Introduction

Financial markets play a crucial role in the global economy, as market transactions incorporate traders' beliefs and expectations regarding future financial conditions. Among the wide range of financial instruments, derivatives stand out due to their diversity in form and purpose. Financial derivatives have been responsible, among other types of contracts, for a significant increase in trading volume.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16760, pp. 149–166, 2027.
https://doi.org/10.1007/978-3-032-30521-3_10

There are several types of derivatives, with futures, forwards, and options being among the most prominent. In particular, option contracts provide valuable information about investors' expectations concerning the price of the underlying asset at the contract's maturity date.

Cox and Ross [14] showed that the value of an option can be expressed as the discounted expected value of its future payoff under the risk-neutral probability distribution of the underlying asset price at maturity. Subsequently, Banz and Miller [7] as well as Breeden and Litzenberger [11], demonstrated that the second derivative of the option pricing function with respect to the strike price is proportional to the risk-neutral density (RND) in an arbitrage-free market.

Due to the richness of the information embedded in option prices, estimating the risk-neutral density at maturity serves several important purposes, including the pricing of illiquid and exotic options, as well as providing insights into investors' risk preferences.

A wide variety of methods has been proposed for estimating RNDs, which can be broadly classified into structural and non-structural approaches. Structural methods assume a stochastic process for the underlying asset and, in some cases, for its volatility (e.g., [9,10,19,24]). In contrast, non-structural methods directly approximate the RND function, potentially involving a large number of parameters. These methods can be further divided into parametric, non-parametric, and semi-nonparametric approaches. Parametric methods rely on a small number of parameters [5,23], whereas non-parametric methods allow for greater flexibility in the shape of the density by using a larger number of parameters [1–3,15,18,33]. Semi-nonparametric methods lie between these two extremes [12]. Some non-parametric approaches rely on estimating the option pricing function and its second derivative [11,14], while others are based on the principle of maximum entropy. Shannon [30] introduced the concept of entropy as a measure of missing information, distinct from the Boltzmann–Gibbs formulation in physics. The maximum entropy principle aims to derive probability distributions from limited information by imposing a set of constraints [20]. In this framework, the maximum entropy density (MED) is considered the "least biased" estimate, as it incorporates all available information without introducing unwarranted assumptions. When applied to financial markets, maximizing entropy leads to the most unbiased probability distribution consistent with observed option prices.

Buchen and Kelly [13] were among the first to apply the maximum entropy principle to recover the RND from European option prices. Stutzer [31] proposed an alternative approach that combines historical data on the underlying asset with entropy maximization to estimate the RND. Further applications include the estimation of volatility surfaces [4], as well as applications to interest rates and bond pricing [8,17].

One of the main objectives of this paper is to propose a measure for the day-to-day evolution of estimated RNDs. This measure is based on optimized entropy values, which arise as a byproduct of the estimation procedures. As discussed above, maximizing entropy subject to constraints derived from option prices allows the estimation of the RND for a given day. In contrast, cross-entropy

provides a measure of the proximity between a given distribution and a prior distribution (see Sect. 2), thus offering an absolute measure of the magnitude of distributional changes between consecutive days. Extensions involving generalized entropy measures, such as Tsallis and Rényi entropies, are explored in [22,32]. Entropy-based optimization methods thus provide a natural framework for constructing measures of the evolution of RNDs over time.

However, for meaningful comparisons over time, a relative measure is required. This is obtained by assessing how much farther the maximum entropy distribution at a given day is from the previous day's distribution, compared to the closest possible distribution consistent with current information. This closest distribution is obtained via minimum cross-entropy.

Rompolis [28], instead of using option prices directly as constraints, estimates the RND based on empirical risk-neutral moments of log-returns derived from option prices. Related formulations for extracting non-central moments are presented in [6,29]. Neri [25] and Schneider [26] estimate maximum entropy distributions using exotic options and apply them to interpolate implied volatilities.

The approach proposed in this paper addresses the issue of ill-conditioning by selecting relevant strike prices based on open interest, which serves as a proxy for the significance of each option contract. A novel empirical feature is identified: the magnitude of open interest influences the selection of relevant strikes and, consequently, the convergence of the estimation methods. This methodology allows focusing on the most informative option prices while mitigating numerical instability, particularly issues related to the ill-conditioning of the Hessian matrix. The quality of the proposed approach is evaluated using both simulated and empirical data. In the case of S&P 500 data, an interesting pattern is observed: open interest exhibits periodic peaks with a cycle of approximately 50.

The remainder of the paper is organized as follows. Section 2 presents the optimization problems associated with maximum entropy and minimum cross-entropy within the Buchen and Kelly framework, as well as the maximum entropy formulation in the Rompolis's approach. Accuracy metrics, including the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), are also introduced. Section 3 applies these methods to simulated data generated from the Black–Scholes and Heston models, as well as to empirical data from the S&P 500, VIX, and Microsoft (MSFT). Measures of RND dynamics are computed and analyzed. Finally, Sect. 4 concludes the paper.

2 Maximum Entropy Density Estimation Methods Using Option Prices

The Principle of Maximum Entropy (PME) allows to select the distribution that maximizes the entropy measure with appropriate constraints.

Consider X a continuous random variable, with probability density function $f(.)$, representing either the price of a security at some future moment T or some price function such as its log-return prices, taking values in an interval I.

According to the maximum entropy principle, the estimation of the density $f(.)$ is obtained from the optimization problem,

$$maximize - \int_I h(f(x))dx \tag{1}$$

$$s.t. \quad \int_I f(x)dx = 1 \tag{2}$$

$$E[h_i(X)] = a_i \tag{3}$$

$$f(x) \geq 0, \quad \forall x \in I \tag{4}$$

where h is a proper choice, in this paper it is considered the Shannon entropy,

$$-\int_I f(x)log(f(x))dx,$$

$I = [0, K[,\ K > 0$, h_i are proper functions.

Buchen and Kelly [13] defined an optimization problem related to the RND estimation obtained from option prices, and based on Shannon entropy as:

$$maximize \quad -\int_0^{+\infty} f(x)log(f(x))dx \tag{5}$$

$$s.t. \quad \int_0^{+\infty} f(x)dx = 1 \tag{6}$$

$$C_i = E[e^{-r\tau}(X - K_i)^+] = \int_0^{+\infty} e^{-r\tau} f(x)(x - K_i)^+ dx,\ i = 1 \ldots, nc \tag{7}$$

$$P_j = E[e^{-r\tau}(K_j - X)^+] = \int_0^{+\infty} e^{-r\tau} f(x)(K_j - x)^+ dx,\ i = 1 \ldots, np \tag{8}$$

where X is the underlying asset price, r is the risk-free interest rate and C_i, $i = 1 \ldots, nc$, and P_j, $j = 1, \ldots, np$, are the derivatives prices, respectively the call and put option prices, with maturity T, $z^+ = max(0, z)$, K_i are the call strike prices, K_j are the put strike prices, nc (resp. np) is the number of calls (resp. puts).

Rompolis [28] optimization problem has the same objective function as in [13], the main difference between them lies in the constraints, [28] uses risk neutral non-central moments, expressed in terms of option prices as constraints. By considering $X = \log\left(\frac{S_T}{S_t}\right)$, where S_t is the underlying asset price, [28] defined an optimization problem related to the RND estimation obtained from option

prices as

$$maximize - \int_D f(x)log(f(x))dx \tag{9}$$

$$s.t. \quad \int_D f(x)dx = 1 \tag{10}$$

$$\int_D x^i f(x)dx = \mu_i,\ i = 1\ldots, m \tag{11}$$

$$f(x) \geq 0, \quad \forall x \in I \tag{12}$$

where $D = \left[\log\left(\frac{K_0}{S_t}\right), \log\left(\frac{K_\infty}{S_t}\right)\right]$, K_0 and K_∞ are the minimum and maximum strike prices given by data, respectively, μ_i denotes the ith order non-central moment of X and $\mu_0 = 1$. From constraint Eq. (11), for $i = 1, 2, 3, 4$, we may obtain the mean, the variance, the skewness, and the kurtosis coefficients, respectively.

Bakshi et al. [6] proposed a problem for recovering RND from option prices through non-central moments, where the risk-neutral moments are represented as follows:

$$\mu_1 = e^{r\tau}\left(1 - \int_{S_t}^{+\infty} \frac{1}{K^2} C_t(\tau, K)dK - \int_0^{S_t} \frac{1}{K^2} P_t(\tau, K)dK\right) - 1$$

$$\mu_i = e^{r\tau}\left(\int_{S_t}^{+\infty} \frac{i}{K^2}\left[log\left(\frac{K}{S_t}\right)^{i-2}\right]\left[i - 1 - log\left(\frac{K}{S_t}\right)\right] C_t(\tau, K)dK + \right. \tag{13}$$

$$\left. \int_0^{S_t} \frac{i}{K^2}\left[log\left(\frac{K}{S_t}\right)\right]^{i-2}\left[i - 1 - log\left(\frac{K}{S_t}\right)\right] P_t(\tau, K)dK\right),$$

where $C_t(\tau, K)$ and $P_t(\tau, K)$ denote the price of a European call and put at time t, respectively.

Rockinger and Jondeau [27] proposed an equivalent unconstrained minimization problem

$$minimize\ Z(\lambda_1, \ldots, \lambda_n) = \int_D exp\left(\sum_{i=1}^{m} \lambda_i(x^i - \mu_i)\right), \tag{14}$$

$\lambda = (\lambda_1, \ldots, \lambda_n)'$ represent the Lagrange multipliers. The solution of the above problem can be rewritten as:

$$f(x) = \frac{1}{Z(\lambda_1, \ldots, \lambda_n)} exp\left(\sum_{i=1}^{m} \lambda_i(x^i - \mu_i)\right). \tag{15}$$

Given up to 4 constraints related to four moments, there is not a closed form density, and therefore problem (9)–(12) is solved only through a numerical method.

The Principle of Minimum Cross-Entropy (PMCE) (Kullback-Leibler) is a further extension of the principle of maximum entropy [13,21]. The cross-entropy is a measure of the distance between two densities f and g, [13]. The PMCE aim is to find a density f which, out of all the densities satisfying the given constraints, is closest to g. The objective function to be minimized is defined by:

$$H_C(f,g) = \int_0^{+\infty} f(x) log\left(\frac{f(x)}{g(x)}\right) \tag{16}$$

where f and g are densities, and in this context g is a prior density. The PME maximizes uncertainty while PMCE minimizes a probabilistic distance to a given distribution. Maximizing entropy is equivalent to minimizing cross-entropy with respect to a uniform distribution.

Buchen and Kelly [13] presented the optimization problem that estimates the RND function from option prices, for a given prior distribution g:

$$\begin{aligned} &minimize - \int_0^{+\infty} f(x) log\left(\frac{f(x)}{g(x)}\right) dx \\ &s.t. \qquad \int_0^{+\infty} f(x)dx = 1 \\ &\qquad E[e^{-r\tau}(X-K_i)^+] = \int_0^{+\infty} f(x)e^{-r\tau}(x-K_i)^+ dx = C_i,\ i = 1\ldots, n_c \\ &\qquad E[e^{-r\tau}(K_j-X)^+] = \int_0^{+\infty} f(x)e^{-r\tau}(K_j-x)^+ dx = P_j,\ i = 1\ldots, n_p \\ &\qquad f(x) \geq 0, \quad \forall x \in [0,+\infty[. \end{aligned} \tag{17}$$

As stated before, we will evaluate the quality of these achievements. One of the purposes of this paper is to overcome the drawback in convergence, pointed out by [13], by choosing as relevant strikes those having high open interests.

Accuracy Metrics

To evaluate the outcome of the models we employ the Mean Absolute Error (MAE) that expresses the average absolute difference between estimates and real values, and Root Mean Square Error (RMSE) that measures the square root of the average squared difference between estimates and real values. Considering $\hat{V}_i$ as estimated prices and V_i as observed prices we have:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{V}_i - V_i)^2}{n}},$$

$$MAE = \frac{\sum_{i=1}^{n}|\hat{V}_i - V_i|}{n}.$$

3 Numerical Experiments

The entropy optimization problems will be tested by using simulated and market option prices. The former are simulated from Black-Scholes-Merton and Heston models. Market option prices on S&P500 and VIX indexes and Microsoft option data are collected from Chicago Board on Options Exchange (CBOE) market.

The optimization problems related to the Problem of Maximum Entropy (PME) are given by (5)–(8) ([13] approach) and (9)–(12) ([28] approach). The optimization problem related to the Principle of Minimum Cross-Entropy problem is given by (17).

Rompolis [28] method was implemented for $m \in \{2, 4\}$ and was designated by numerical maximum entropy density estimation based on risk-neutral moments (MED-RNM-N). For $m = 2$ we can obtain the analytical solution of the optimization problem, therefore, this approach is denoted by analytical maximum entropy density estimation based on risk-neutral moments (MED-RNM-A). Since the approach for $m = 4$ gives poor results, we choose to compare the analytical solution with the numerical solution for $m = 2$.

A simulation analysis for BSM and Heston data is performed using the routines *blsprice* and *optByHestonNI*, respectively, from Financial Toolbox of Matlab. For implementing Buchen and Kelly's method we used the CVX routine which is a matlab-based program for convex optimization, complemented with new built functions.

3.1 Simulation Data Black-Scholes Model

The stock price model proposed by [10] and [24] assume that the underlying asset follows a geometric Brownian motion,

$$dS_t = S_t \mu dt + S_t \sigma dW_t,$$

where μ and σ are constants and denote the expected return and the volatility, respectively, S_t is the underlying asset price at time t and W_t a standard Brownian motion. Applying Itô Lemma allows to specify that the probability distribution of $log(S_T)$, $T > t$, is given by

$$log(S_T) \sim N\left(log(S_t) + (\mu - \frac{1}{2}\sigma^2)\tau, \sigma^2\tau\right),$$

where $\tau = T - t$ and $N(\alpha, \beta^2)$ denotes a normal distribution with mean α and variance β^2. Therefore S_T follows a log-normal distribution.

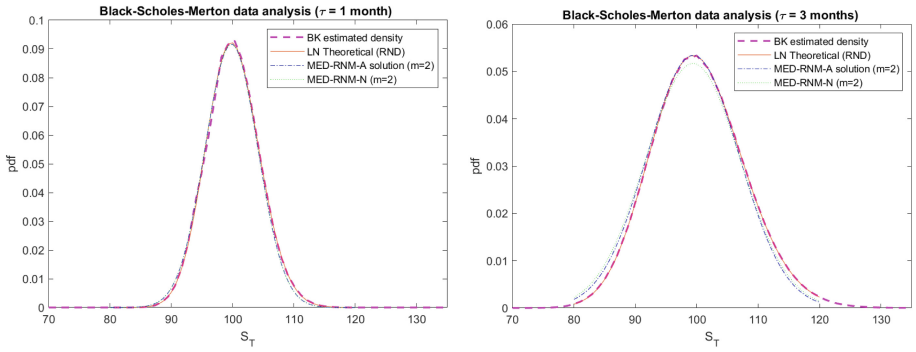
In order to generate BSM option prices, the chosen model parameters are: the current underlying asset price, $S_t = 100$, the risk-free interest rate, $r = 0.01$, which is the return under risk neutral distribution and the volatility, $\sigma = 0.15$. We considered 21 strike prices, K, equally spaced between 80 and 120, and two times to maturity $\tau \in \{\frac{1}{12}, \frac{3}{12}\}$. Therefore, we generated 42 option prices (21 calls and 21 puts), for each maturity.

Due to different liquidity of the option contracts and transaction costs, the simulated call prices were perturbed. To reproduce market behavior, random noise is added to the simulated prices:

$$C_i^+ = (1 + 0.005\xi_i)C_i \tag{18}$$

$$P_i^+ = (1 + 0.005\eta_i)P_i \tag{19}$$

$\xi \sim N(0,1)$, $\eta \sim N(0,1)$, and C_i, P_i are the theoretical call and put prices computed by BSM formula. The multiplicative form of (18)–(19) allows to mimic market heterocedasticity.



(a) (b)

Fig. 1. Theoretical (lognormal) RND and estimated RNDs from BSM data - Buchen and Kelly, and Rompolis methods - 1 and 3 months

Table 1. RMSE and MAE for estimated call and put option prices (BSM data)

	Buchen and Kelly method	MDN-RNM-A method	MDN-RNM-N method (m = 2)
RMSE (one month)	3.2695×10^{-4}	1.2×10^{-2}	1.3×10^{-2}
RMSE (three months)	1.5307×10^{-4}	8.8717×10^{-2}	8.8749×10^{-2}
MAE (one month)	1.0596×10^{-4}	8.817×10^{-3}	8.817×10^{-3}
MAE (three months)	4.1365×10^{-5}	1.3×10^{-2}	1.6×10^{-2}

By analyzing the RMSE and MAE values in Table 1, we observe that the estimated values are accurate, which supports the validity of the proposed estimation methods for theoretical data. The results appear to be more accurate for Buchen and Kelly's method than for Rompolis's method. In Fig. 1, the theoretical and estimated densities are plotted together. Overall, the estimation

shows a good fit, with better performance for Buchen and Kelly's method. However, some issues arise with the Rompolis's method. We observe that empirical RND moments may not capture all the relevant information contained in option prices. In particular, different behaviors are observed across distinct regions of the density support. In this regard, Buchen and Kelly's method is more flexible, as it captures different behaviors between each pair of consecutive strikes.

3.2 Heston Data Analysis

The effectiveness of the Buchen and Kelly's method is also analyzed through estimation of RNDs for Heston model, [19]. He assumed that volatility varies stochastically. The dynamic model has two stochastic components, the underlying asset price S_t and its variance V_t,

$$dS_t = S_t \mu dt + S_t \sqrt{V_t} dWt, 1, \tag{20}$$

$$dV_t = \mathcal{K}(\theta - V_t)dt + \sigma \sqrt{V_t} dWt, 2, \tag{21}$$

where $dW_{t,1}$ and $dW_{t,2}$ are standard Brownian motions with correlation coefficient ρ, $\mathcal{K}$ represents the volatility mean reversion speed for V_t, θ denotes the long-term mean variance, and σ is the volatility of the volatility process.

The chosen parameters values are: risk-free interest rate, $r = 0.05$; initial variance for the underlying asset, $V_t = 0.026$; volatility mean reversion speed for V_t, $\mathcal{K} = 3.93$; long-term mean of variance, $\theta = 0.01$; volatility of volatility $\sigma = 0.04$; the maturity, $\tau = \frac{1}{12}$ and the underlying asset price, $S_t = 1080$. Also, it is considered 12 strike prices equally spaced between 820 and 1080, corresponding to 12 calls and 12 put prices.

Different correlation coefficients are considered, $\rho \in \{-0.52, 0, 0.52\}$. To reproduce the market's behavior, random noise is added to theoretical option prices. Hence, we replaced these prices as follows, $C_i^* = (1 + 0.005 \cdot \xi_i) \cdot C_i$ and $P_j^* = (1 + 0.005 \cdot \eta_j) \cdot P_j$, with $\xi \sim \mathcal{N}(0, 1)$ and $\eta \sim \mathcal{N}(0, 1)$.

Table 2 compares theoretical and estimated densities, it is observed that for all ρ values, the estimation is overall well fitted. Nonetheless, the obtained estimates present worst adjustment than for BSM data (though not so clear for MAE measure), even without noise. Figure 2 presents that for $\rho = 0.52$ the skewness is positive, which means that the right tail is heavier, it shows also that by perturbing option prices, we get worse estimates.

Table 2. RMSE and MAE for estimated call and put option prices (Heston data without noise)

ρ	RMSE	MAE
-0.52	1.1×10^{-2}	7.6860×10^{-4}
0	1.1×10^{-2}	7.6434×10^{-4}
0.52	1.3×10^{-2}	7.7524×10^{-4}

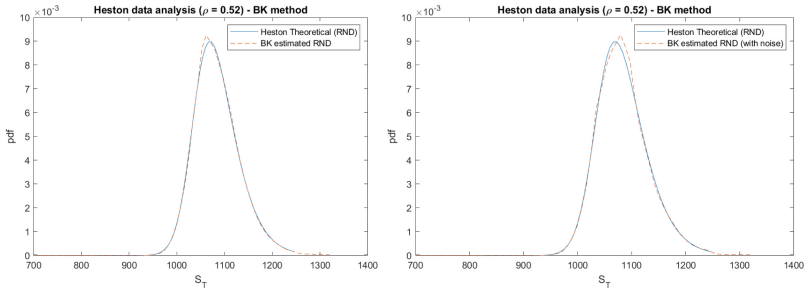


Fig. 2. Heston theoretical RND and estimated RNDs from Heston data (with and without noise)

3.3 Market Data Analysis

Risk neutral densities are recovered using option prices from $S\&P500$ and VIX indexes and from Microsoft data. These data were obtained from CBOE exchange market. We choose Buchen and Kelly's method to evaluate the densities.

Dataset Description. For each sample, we considered the arithmetic mean of bid and ask option prices. The interest rate data is collected from the Central Bank of the Federal Reserve System of the United States of America. To ensure the quality of the data, we selected call and put option prices with larger open interests since they can possibly better represent the market's trader expectations. For $S\&P500$ data, we consider a sequence of strikes with a constant difference of 50 between consecutive ones, since these were the ones with remarkably higher open interest values, for the period considered. When considering other strike prices, we didn't obtain better results and in some cases not even convergence. We stress that this is a feature that strongly improves the quality and feasibility of Buchen and Kelly's estimation.

We consider three datasets from the $S\&P500$ index. The first dataset collects data from November 1st to 12th, 2021, with maturity on December 17th, 2021. The second dataset collects data from January 10th to 14th, 2022, with maturity on February 18th, 2022, this dataset also contains data from VIX from January 10th to 14th, 2022, with maturity on February 16th, 2022. The third dataset collects data from November 1st, 2021, with maturity on January 21st, 2022, $S_t = 4614$, and $r = 0.05\%$. We estimated maximum entropy densities for all datasets and minimum cross-entropy density for the first dataset. Table 3 present some information about first and second datasets regarding S&P500 and VIX, respectively.

Microsoft dataset includes data from December 1st, 2021, with maturity on June 17th, 2022. The underlying asset price, S_t, is 330 and the interest rate, r, is 0.001. Since Microsoft option data are American style options, we selected only out-of-the money options.

Table 3. Datasets

First dataset - $S\&P500$				Second dataset - VIX			
Day	τ (days)	S_t	$r(\%)$	Day	τ (days)	S_t	$r(\%)$
November 1st	46	4613	0.08	January 10th	37	19.4	0.05
November 2nd	45	4630	0.06	January 11th	36	18.41	0.04
November 3rd	44	4660	0.07	January 12th	35	17.62	0.04
November 4th	43	4680	0.05	January 13th	34	20.31	0.05
November 5th	42	4697	0.06	January 14th	33	19.19	0.05
November 8th	39	4701	0.05				
November 9th	38	4685	0.05				
November 10th	37	4646	0.06				
November 11th	36	4649	0.06				
November 12th	35	4682	0.05				

Estimated RND Results. Considering the first dataset for S&P500 data, Figs. 3, 4 present the estimated daily RND.

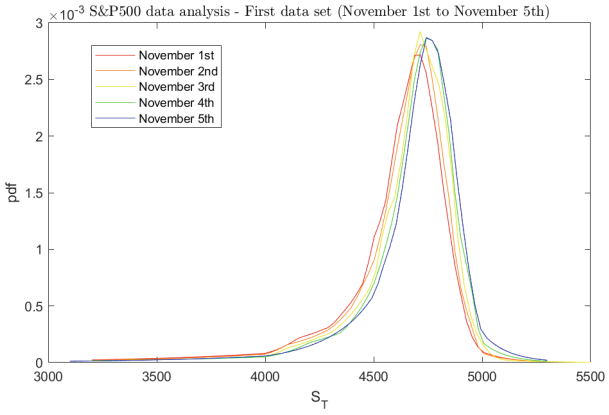


Fig. 3. Estimated RNDs from $S\&P500$ (first dataset) - November 1st to November 5th

As for Table 4, we present the statistical parameters for estimated RNDs. The mean for the density estimates is situated between 4604 and 4693. Remark that, in general, the mean values are near the current value (Table 3).

Considering Figs. 3, 4, and Table 4 we observe that skewness is negative, which means that left tail is heavier, and therefore the probability of strongly negative returns is higher than for positive returns. The kurtosis coefficient is greater than 3 (leptokurtic distribution), that suggest there may be a large probability of extreme values occurrence on both sides.

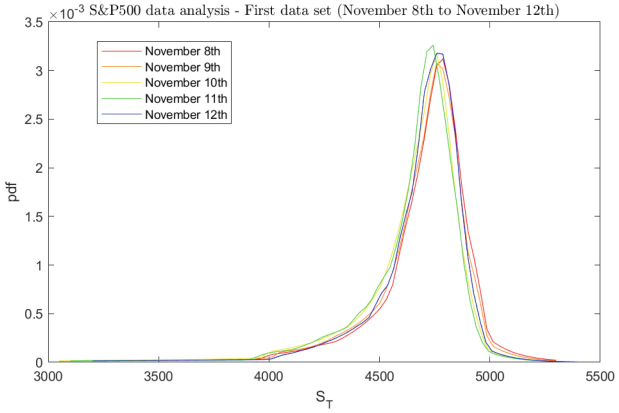


Fig. 4. Estimated RNDs from $S\&P500$ (first dataset) - November 8th to November 12th

Table 4. Statistical parameters for RND estimate (first dataset) - $S\&P500$

Day	Mean	Variance	Skewness	Kurtosis
November 1st	4604	6832	−2.0395	9.0948
November 2nd	4621	6467	−2.1058	9.5512
November 3rd	4651	5835	−2.1755	10.5705
November 4th	4672	5794	−2.2311	11.0139
November 5th	4685	6332	−2.2628	11.2655
November 8th	4693	6230	−2.4338	12.7395
November 9th	4676	6308	−2.3495	11.8706
November 10th	4642	6567	−2.3242	11.3210
November 11th	4645	5754	−2.3119	11.7340
November 12th	4682	4973	−2.3259	12.2705

There seems to be a relationship between skewness and kurtosis. When skewness goes up (resp. down) in absolute value, so does the kurtosis.

Considering the estimation applied to $S\&P500$ data for the second dataset, Table 5, we observe that as in the previous case, the skewness is negative, and the skewness and kurtosis increase when the mean increases. Negative skewness may indicate that investors are somewhat pessimistic.

As for the estimation applied to VIX data, Table 6 and Fig. 5, we observe that the asymmetry is positive. There seems to be a nearly inverse relationship between the asymmetry obtained for the $S\&P500$ and for the VIX, see Tables 5 and 6. When skewness increases in absolute value for $S\&P500$ data, it decreases for VIX data, and vice versa, in general, as it is expected. In addition, we can state that for $S\&P500$ data, the main interest is the estimation of the left tail

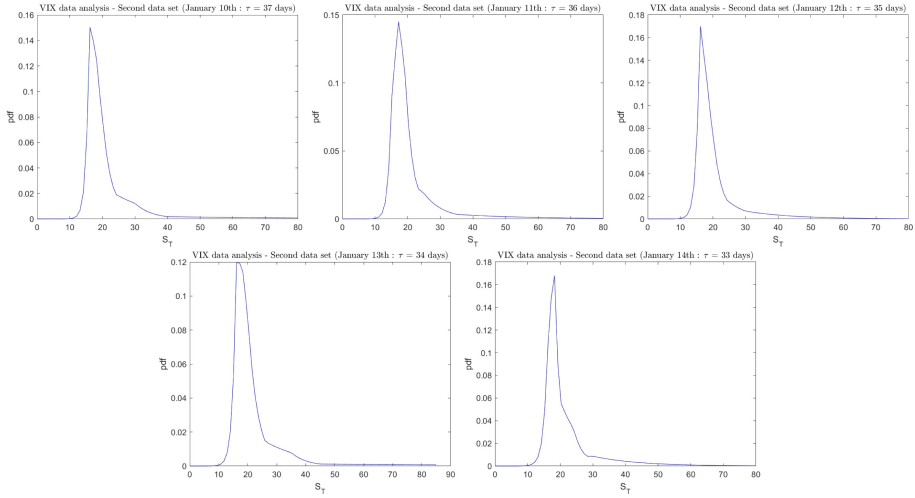


Fig. 5. Estimated RNDs from VIX (second dataset)

Table 5. Statistical parameters for estimated RNDs (second dataset) - $S\&P500$

Day	Mean	Variance	Skewness	Kurtosis
January 10th	4666	7417	−1.9786	8.9133
January 11th	4709	6765	−2.0481	9.4916
January 12th	4725	6179	−2.0719	9.8647
January 13th	4656	7464	−1.8074	7.6025
January 14th	4663	6584	−1.8186	7.9930

associated with more risk, whereas, for VIX data, it is the right one associated with more volatility, see Figure 5.

Considering Fig. 6 and Table 7, S&P500 - third dataset, we observe that just as in the first dataset, the skewness is negative, and the kurtosis is high. As for Fig. 7 and Table 7, MSFT, we observe that just as for the previous estimated RND, the skewness is negative, and the kurtosis is high, but both seem to be more limited than for S&P500.

Comparison Between Minimum Cross-Entropy and Maximum Entropy Densities. In this paper it is proposed an indicator that allows us to compare each RND estimate to that of the previous day, analyzing if there was a substantial variation. This indicator, $M(H_C)$, is defined as follows:

$$M(H_C)_t = \frac{H_C(f_t, f_{t-1})}{H_{min}} = 1 + \frac{H_C(f_t, f_{t-1}) - H_{min}}{H_{min}} \tag{22}$$

Table 6. Statistical parameters for estimated RNDs - VIX

Day	Mean	Variance	Skewness	Kurtosis
January 10th	21.68	91	3.2825	15.4297
January 11th	21.10	84	3.2474	15.3343
January 12th	20.85	72	3.0919	14.4984
January 13th	22.15	94	3.3166	16.1484
January 14th	21.85	80	2.8095	12.4113

Table 7. Statistical parameters for estimated RND (third dataset) - $S\&P500$ and $MSFT$

	Mean	Variance	Skewness	Kurtosis
Third dataset- $S\&P500$	4661	1265	−1.7981	7.2333
$MSFT$	330	5553	−0.3525	4.2285

where f_t (resp. f_{t-1}) represents the maximum entropy density, estimated at day t (resp. $t-1$); $H(f_t, f_{t-1})$ denotes the cross-entropy between f_t and f_{t-1}; H_{min} denotes the cross-entropy between f_t^* and f_{t-1} defining f_t^* as the density closest to f_{t-1}, given time t, that is, the density obtained by fitting the minimum cross-entropy with respect to f_{t-1}, subject to the constraints at time t.

If $M(H_C)$ is very close to 1, there is little change with respect to the estimated density of the previous day since the estimated density for day t does not have a cross-entropy with day $t-1$ density much higher than the possible minimum one. Therefore $M(H_C)$ measure can serve as an effective diagnostic tool for identifying days with the most significant changes in the market sentiment, allowing for the subsequent identification of explanations based, for example, on unexpected events. It is also worth examining which density zones exhibit the greatest divergence between H_C and H_{min} on such days.

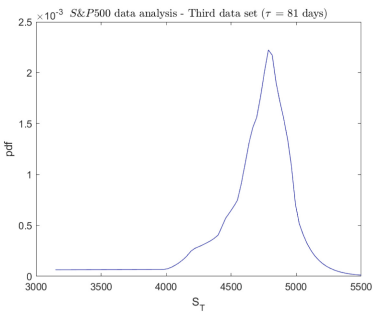


Fig. 6. Estimated RND from $S\&P500$ (third dataset)

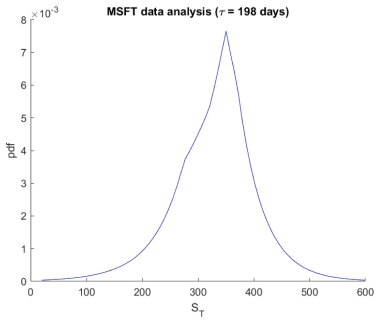


Fig. 7. Estimated RNDs from $MSFT$

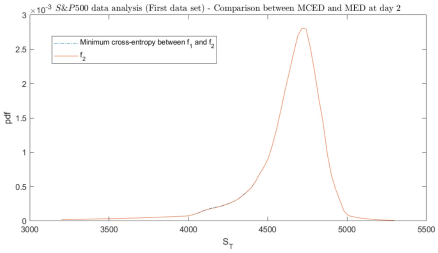


Fig. 8. $S\&P500$ data analysis (First dataset) - Comparison between MCED and MED at day November 2nd.

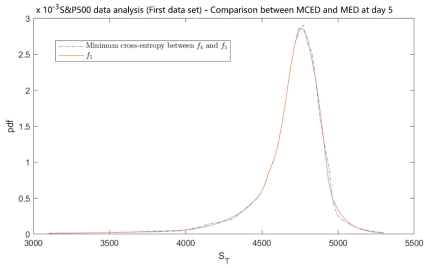


Fig. 9. $S\&P500$ (First dataset) - Comparison between MCED and MED at day November 5th

Considering Fig. 8, it is illustrated the comparison between the minimum cross-entropy density (MCED) and the maximum entropy density (MED) estimates, for the first dataset, November 2nd. The estimates obtained for MCED are similar to those obtained for MED, this may signify that there is little difference between the current density and the density estimate of the previous day.

Through Table 8 it is possible to perceive that the highest ratio, $M(H_C)$, was on November 11th, followed by November 5th. Figure 9, corresponding to November 5th, shows that MCED estimate does not adjust so well to MED estimate as for November 2nd (Fig. 8). It seems that the ratio reflects well changes in the estimated densities, over time. It would be interesting to test the behavior of the ratio in case of more drastic evolution in densities. The highest H_{min} value was for November 4th and the lowest value for November 9th.

Table 8. Indicator $M(H_C)$

	H_C	H_{min}	$M(H_C)$
November 2nd (prior: November 1st)	0.0055	0.0054	1.0031
November 3rd (prior: November 2nd)	0.2851	0.2847	1.0012
November 4th (prior: November 3rd)	0.7332	0.7331	1.0012
November 5th (prior: November 4th)	0.0359	0.0354	1.0673
November 8th (prior: November 5th)	0.0119	0.0119	1.0057
November 9th (prior: November 8th)	0.0051	0.0050	1.0260
November 10th (prior: November 9th)	0.0211	0.0208	1.0179
November 11th (prior: November 10th)	0.0146	0.0134	1.0934
November 12th (prior: November 11th)	0.1493	0.1486	1.0044

4 Conclusions

Option prices can provide relevant information to evaluate the traders' expectations and risk preferences. From option data, it is then possible to estimate the RND, which could reflect the market sentiment that traders attribute to the future price of the underlying asset. This work applies entropy optimization methods based on PME, Buchen and Kelly and Rompolis methods, and on PMCE for Buchen and Kelly method. For this purpose, we used simulated data from BSM and Heston models, and market data on S&P500, VIX, and MSFT.

Furthermore, innovations were developed for implementing these methods for market data. First, a selection of strikes was performed in order to operationalize the Buchen and Kelly method. Second, a new measure was conceived to analyse a short period of time, and perceive if there was a substantial variation in RND at each day regarding the previous day.

Through graphical analysis for the simulated data, we highlight that both methods produce valid risk-neutral density estimates. We obtained proper RND estimates, always with non-negative values and with expected tail behavior. Regarding the market data, the estimated RNDs verify the expected behavior. It is possible to observe for S&P500 and MSFT datasets that investors seem to be pessimistic about the market since skewness is negative. In addition, the kurtosis is high, which may signify that the risk is high. Regarding VIX data, the skewness is positive in opposite to the results from S&P500 and Microsoft. The right tail is the main interest of the estimation since it can reflects traders' concerns with high price flutuations, namely the negative ones.

Acknowledgments. This work has been funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., through project UID/05037/2025 with DOI https://doi.org/10.54499/UID/05037/2025, and through the project "Complex Risk Management in the Big Data Regime" – PTDC/MAT–APL/1286/2021.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aït-Sahalia, Y., Duarte, J.: Nonparametric option pricing under shape restrictions. J. Econometrics **116**(1), 9–47 (2003)
2. Aït-Sahalia, Y., Lo, A.W.: Nonparametric estimation of state-price densities implicit in financial asset prices. J. Financ. **53**(2), 499–547 (1998)
3. Ardakani, O.M.: Option pricing with maximum entropy densities: the inclusion of higher-order moments. J. Futur. Mark. **42**(10), 1821–1836 (2022)
4. Avellaneda, M.: Minimum-relative-entropy calibration of asset-pricing models. Int. J. Theor. Appl. Financ. **1**(04), 447–472 (1998)
5. Bahra, B.: Implied risk-neutral probability density functions from option prices: theory and application. Bank of England Working Paper (1997)
6. Bakshi, G., Kapadia, N., Madan, D.: Stock return characteristics, skew laws, and the differential pricing of individual equity options. Rev. Financ. Stud. **16**(1), 101–143 (2003)

7. Banz, R.W., Miller, M.H.: Prices for state-contingent claims: some estimates and applications. J. Bus. 653–672 (1978)
8. Bariviera, A.F., Martín, M.T., Plastino, A., Vampa, V.: Libor troubles: anomalous movements detection based on maximum entropy. Phys. A **449**, 401–407 (2016)
9. Bates, D.S.: Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. Rev. Financ. Stud. **9**(1), 69–107 (1996)
10. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**(3), 637–654 (1973)
11. Breeden, D.T., Litzenberger, R.H.: Prices of state-contingent claims implicit in option prices. J. Bus. 621–651 (1978)
12. Bu, R., Hadri, K.: Estimating option implied risk-neutral densities using spline and hypergeometric functions. Economet. J. **10**(2), 216–244 (2007)
13. Buchen, P.W., Kelly, M.: The maximum entropy distribution of an asset inferred from option prices. J. Financ. Quant. Anal. **31**(1), 143–159 (1996)
14. Cox, J.C., Ross, S.A.: The valuation of options for alternative stochastic processes. J. Financ. Econ. **3**(1–2), 145–166 (1976)
15. Dalderop, J.: Nonparametric filtering of conditional state-price densities. J. Econometrics **214**(1), 295–325 (2020)
16. Fernandes, A.: Analysis of investors expectations using entropy optimization models. Master's thesis, University of Coimbra (2022). https://estudogeral.uc.pt/handle/10316/100098
17. Gzyl, H., Mayoral, S.: Determination of zero-coupon and spot rates from treasury data by maximum entropy methods. Phys. A **456**, 38–50 (2016)
18. Härdle, W., Hlávka, Z.: Dynamics of state price densities. J. Econometrics **150**(1), 1–15 (2009)
19. Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)
20. Jaynes, E.T.: Concentration of distributions at entropy maxima. ET Jaynes: Papers on Probability, Statistics and Statistical Physics p. 315 (1979)
21. Kullback, S., Leibler, R.: On information and sufficiency. Ann. Math. Stat. **22**, 79–86 (1951)
22. Malhotra, G., Srivastava, R., Taneja, H.: Calibration of the risk-neutral density function by maximization of a two-parameter entropy. Phys. A **513**, 45–54 (2019)
23. Melick, W.R., Thomas, C.P.: Recovering an asset's implied pdf from option prices: an application to crude oil during the gulf crisis. J. Financ. Quant. Anal. **32**(1), 91–115 (1997)
24. Merton, R.C.: Theory of rational option pricing. Bell J. Econ. Manage. Sci. 141–183 (1973)
25. Neri, C., Schneider, L.: Maximum entropy distributions inferred from option portfolios on an asset. Finance Stochast. **16**, 293–318 (2012)
26. Neri, C., Schneider, L.: A family of maximum entropy densities matching call option prices. Appl. Math. Financ. **20**(6), 548–577 (2013)
27. Rockinger, M., Jondeau, E.: Entropy densities with an application to autoregressive conditional skewness and kurtosis. J. Econometrics **106**(1), 119–142 (2002)
28. Rompolis, L.S.: Retrieving risk neutral densities from European option prices based on the principle of maximum entropy. J. Empir. Financ. **17**(5), 918–937 (2010)
29. Rompolis, L.S., Tzavalis, E.: Recovering risk neutral densities from option prices: a new approach. J. Financ. Quant. Anal. **43**(4), 1037–1053 (2008)
30. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Technical J. **27**(3), 379–423 (1948)

31. Stutzer, M.: A simple nonparametric approach to derivative security valuation. J. Financ. **51**(5), 1633–1652 (1996)
32. Tanak, A.K., Borzadaran, G.M., Ahmadi, J.: Maximum Tsallis entropy with generalized Gini and Gini mean difference indices constraints. Phys. A **471**, 554–560 (2017)
33. Yatchew, A., Härdle, W.: Nonparametric state price density estimation using constrained least squares and the bootstrap. J. Econometrics **133**(2), 579–599 (2006)